DØ Computing and Analysis

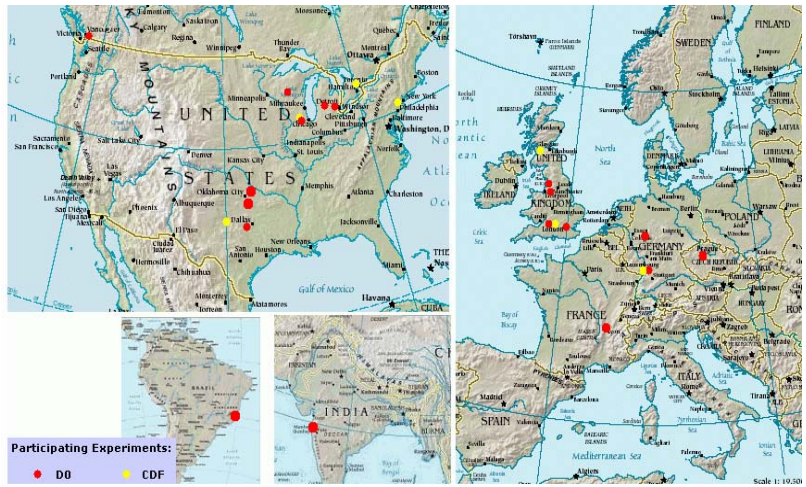
Brad Abbott

DOE Tevatron Operations Review Preparation

March 27, 2007

DØ computing model

- Distributed computing, moving toward automated use of common tools on grid
- Scalable
- Work with OSG/LCG, not against, increased resources
- Need to conform to standards
- DØ running experiment and is taking data. Need to take prudent approach to computing

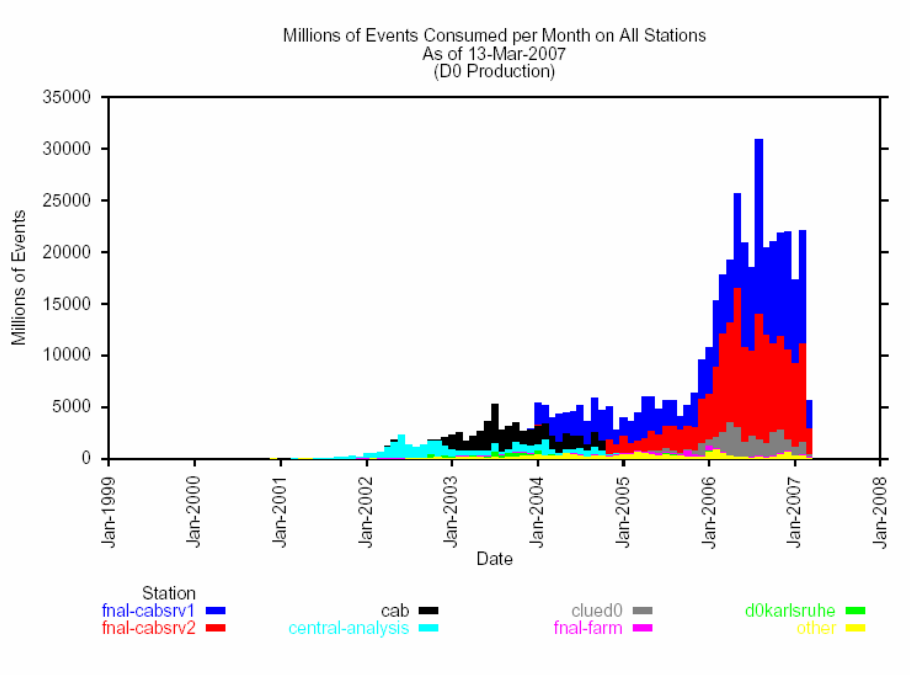


DØ computing model

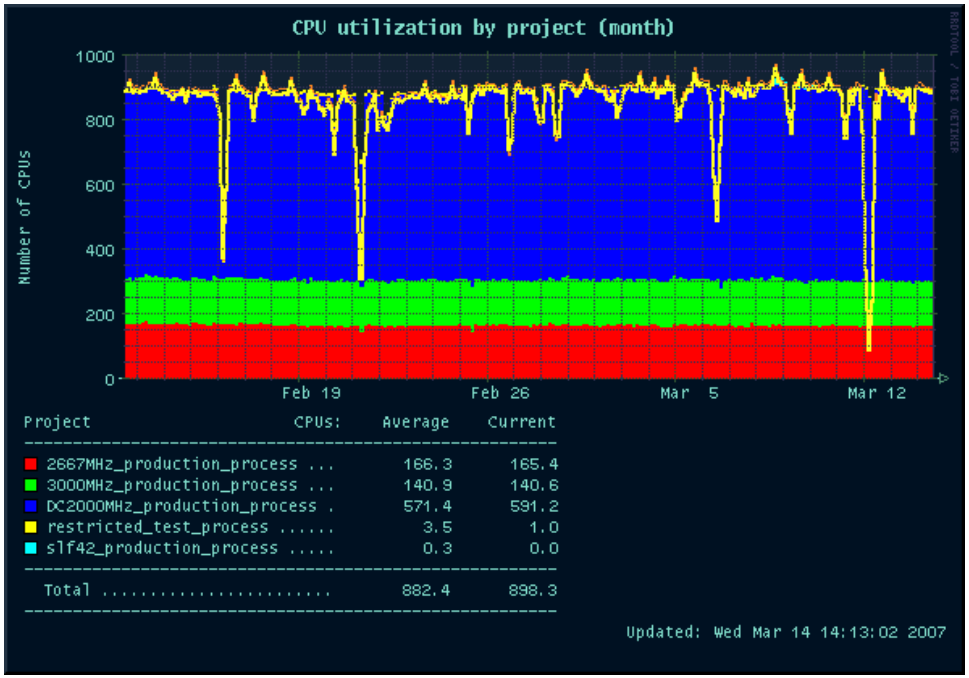
- Started with distributed computing with evolution to automated use of common tools/solutions on the grid (SAMGrid) for all tasks enabling physics analysis
- SAM (Sequential Access to Metadata) used for data handling and storing
- Increased interoperability / automation ⇨ increased production rates along with increased functionality
    - Primary processing with SAMGrid
    - Deployment of SAMGrid v7 (inc. d0runjob, MC, fixing etc)
    - SAM and SAMGrid performance – inc data handling and deployment/operations
    - Full interoperability with OSG/LCG(for Monte Carlo production and data reprocessing)
    - Automated submission of all (Monte Carlo) production jobs by SAM shifters. (Issues with having generic shifters obtaining proper credentials)

Analysis over time

- Number of events processed since "beginning of SAM time"
- Integrates to 450 billion events
- SAM allows a very efficient way to store and access data
- Smoothly operating for many years
- CAB (Central Analysis Backend) – two clusters cabsrv1,cabsrv2: ~ 1000 cpu cluster which is the primary computing resource used for analysis
- Significant computing resources available



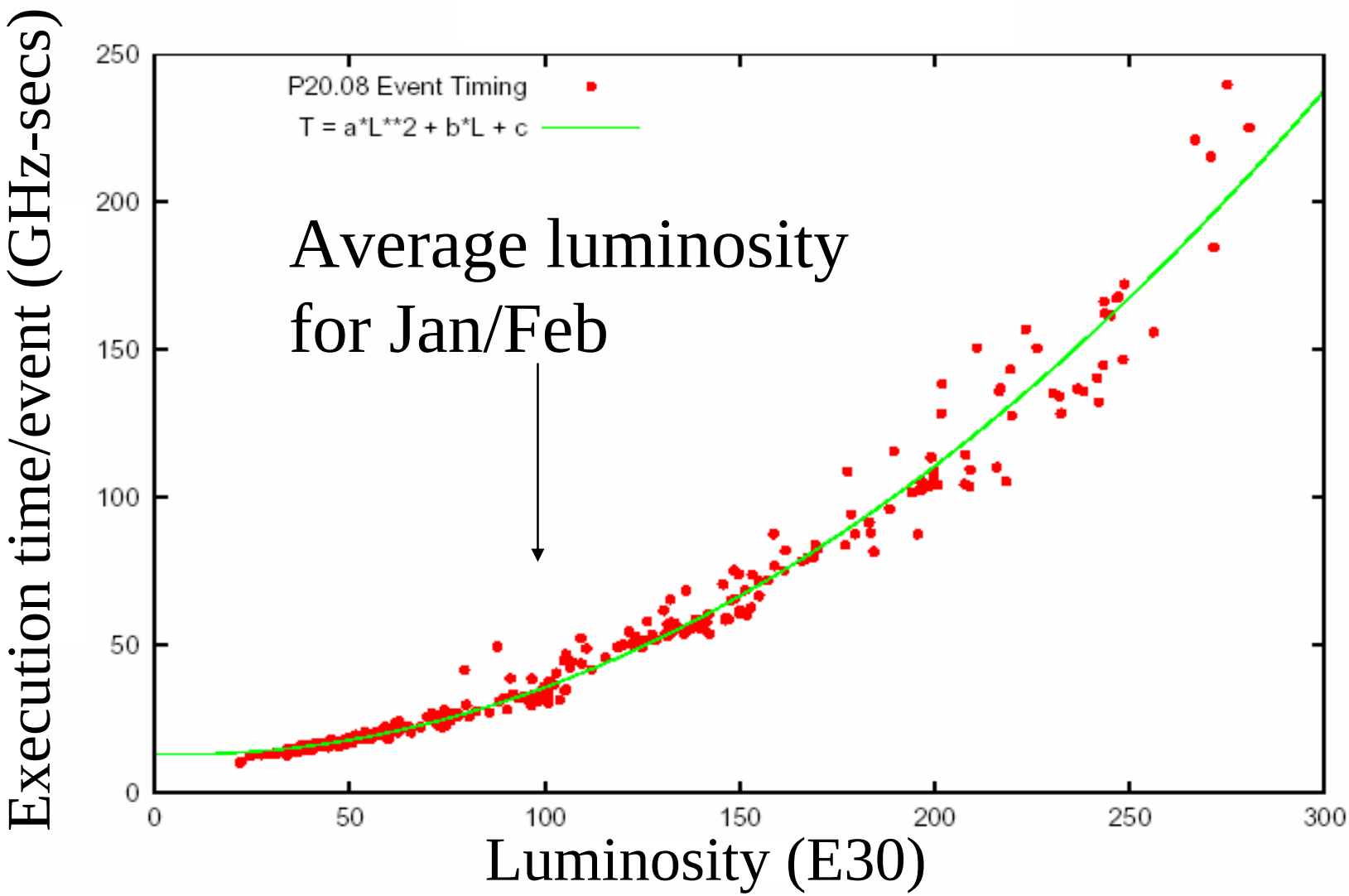
FNAL production farm very efficiently running



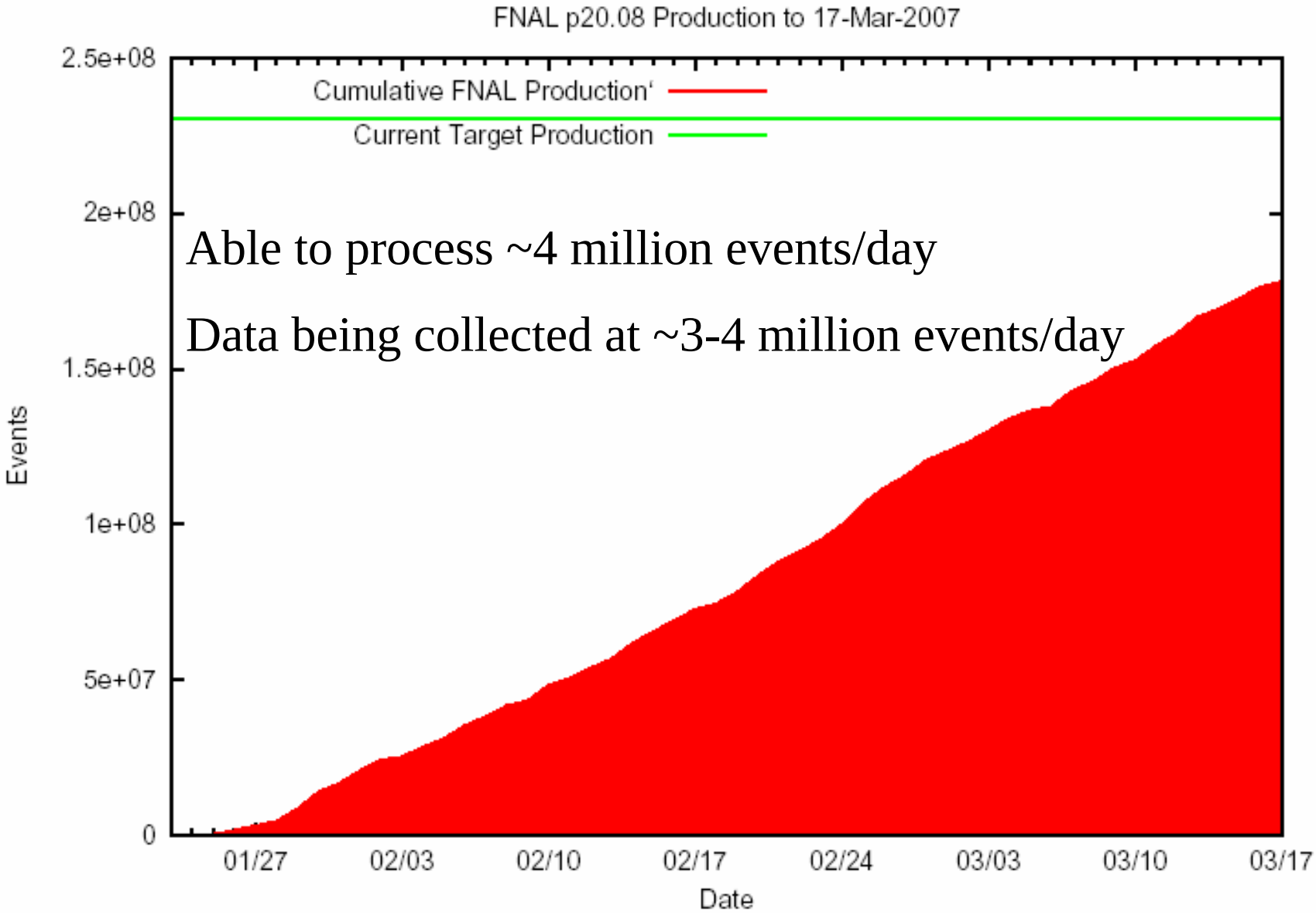
- Running at ~80% efficiency
- Loss of efficiency due to
  - Scheduled downtimes
  - Nodes under repair
  - Occasional power failure
  - …

Node	# nodes	#processes	Normalization	PIII eq.
SC2667	102	204	2.02	412
SC3000	70	140	2.23	312
DC1776	160	640	2.53	1619

Event reconstruction time



Current FNAL Farm Production



Dealing with Higher Luminosities

- 9 racks of additional nodes were purchased for use on FARM and CAB
- Nodes: Intel Xeon 5148, 2.33 GHZ, dual cpu, dual core in each cpu
- 5 racks for FNAL farm, 4 racks CAB.
- Approximately double FNAL farm performance
- We can switch nodes between CAB/FARM depending on needs.
- With addition of new nodes we can continue to process data at current rate even if every store has a peak store luminosity of 300 E 30.
- New disks: 18 satabeasts with 16TB of disk space each. Increased disk space for project disks and sam cache

MC Production Numbers

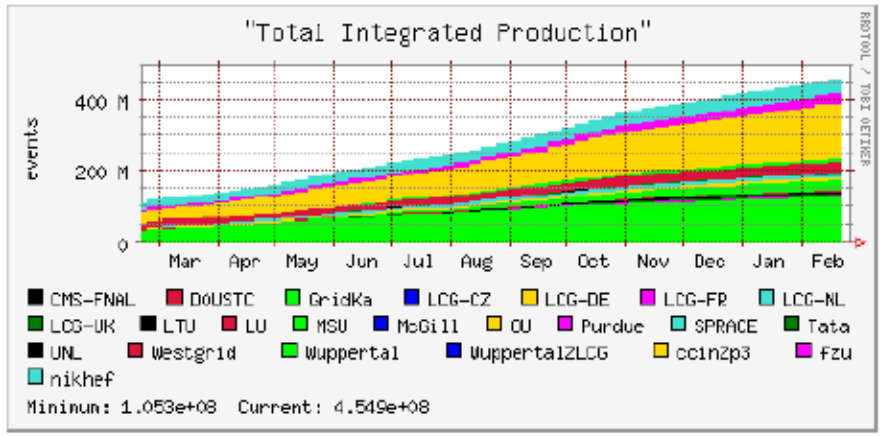
P17 started Sept. 2005

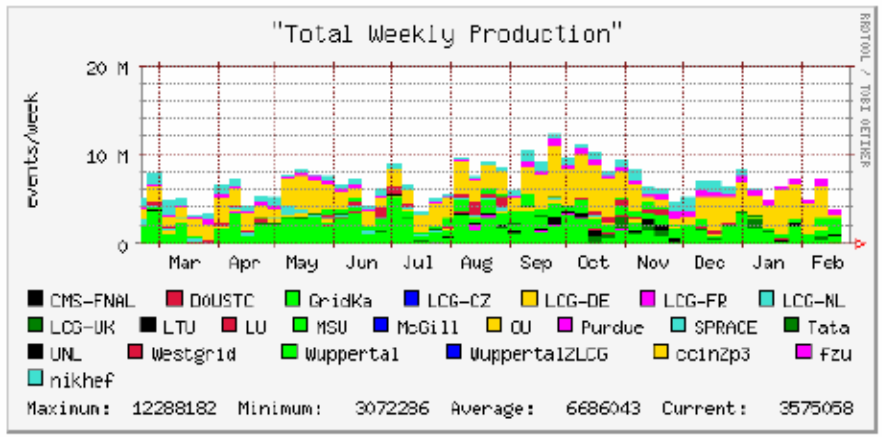
455 M total events
up to February 27

29 M/month last year

57% via Samgrid
2.0% LCG
0.6% OSG

80.1% Europe
17.1% N. America
2.2% S. America
0.6% Asia





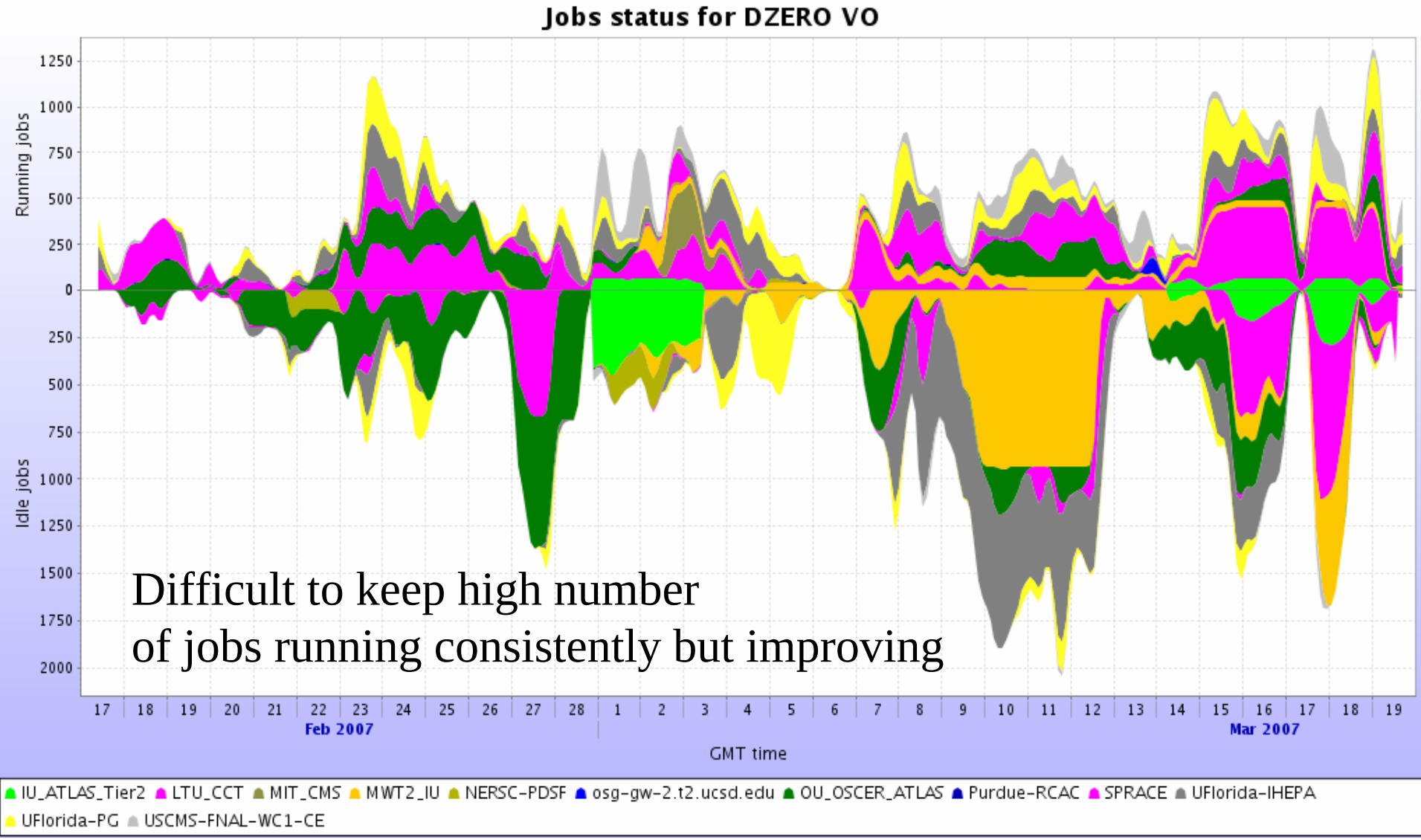
Recent algorithm improvements

- New calorimeter calibration for both EM and HAD calorimeters
- Tracking P_T threshold increased to 450 MeV
- Removed cut on the maximum number of SMT hits at high luminosity
- Removed CFT ADC count cut from reco code set CFT thresholds in CFT database (much lower than old offline cut)
- Lorentz shift correction for layer 0
- ICD calibrations
- Luminosity DB access from D0reco
- Luminosity tick information into the event and stored in TMB
- New version of muon T0's
- Much more robust, very few crashes
- Removed occupancy dependent P_T thresholds in tracking
- Improved tracker alignment

Reprocessing

- Due to improvements in reconstruction code and improvements in detector calibrations, Run II b data taken in 2006 and early 2007 needs to be reprocessed.
- Run IIb data taken in 2006 is ~ 500 million events
- Previously (2004 reprocessing) had large farms with experts running jobs. Currently running on smaller OSG sites and 2 non-OSG sites.
- Shifters (non-experts) running production at multiple OSG sites
- Using 1500-2000 CPUs, trying to finish in 2-3 months

Reprocessing status with OSG



Difficult to keep high number
of jobs running consistently but improving

Future improvements

- Very challenging to get a full production system running for many months at high efficiency on the grid.
- Working with OSG to improve efficiency
  - OSG trouble shooting team working with DØ has improved efficiency
- Bringing in LCG sites to increase CPU
- Will use spare cycles on the upgraded FNAL farm (and temporarily use new CAB nodes) to help with reprocessing
- Even with LHC using a lot of resources we are able to get necessary CPU using opportunistic computing

Analysis on Grid

- Analyzer needed 300,000 CPU hours to process a matrix element top mass analysis. Would have required all of CAB for more than a month. Went to OSG and was able to finish project.
- Another is analyzer running analysis on LCG-grid
- Analyzers are starting to use grid for physics analysis.

Streamlining projects

- Available manpower will be reduced in future so need to automate tasks
- Currently some computing projects are performed manually, i.e. skimming of the data
- Computing planning board has developed a document describing our needs to put many projects into production, thus reducing the need for individuals to run projects
- Reducing connection between runjob(configures and runs DØ executables) and SAMGrid(data handling) to allow more efficient method to introduce new features and to reduce maintenance by SAMGrid team.
- SAM shifters take over responsibility of MC running.
- Studying jobs on CAB to improve CPU efficiency of our analysis jobs

Security

- DØ operates its computing systems in accordance with the Fermilab Policy on Computing and in accordance with guidelines established by the Fermilab Computer Security Team
- Some offline systems are constrained beyond the baseline requirements:
  - Linux desktops run local firewalls
  - Production farms grant access only to limited set of users and that access must come from onsite nodes
- Since Jan 1st, 2006 there have been 2 reportable computer security incidents involving DØ systems
  - No data were lost, no physics results were delayed, nor was there any impact on DØ Online Major Application as a result of either incident

High Luminosity workshop

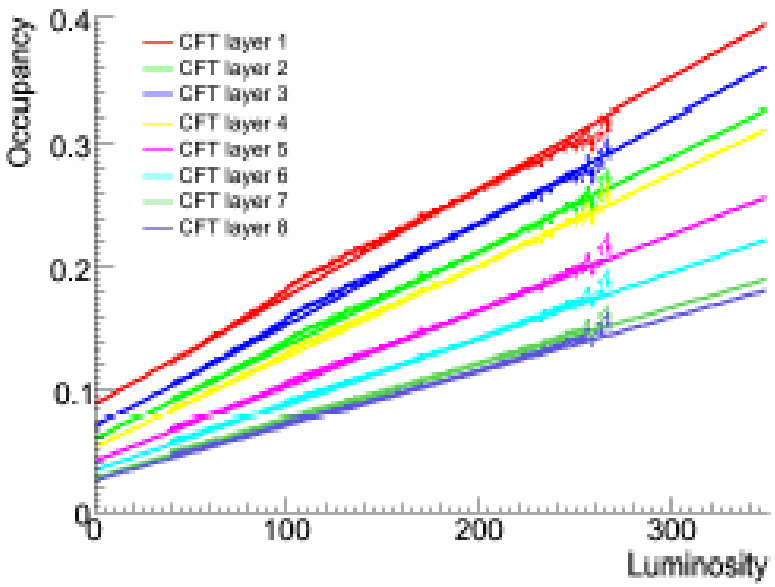
- Tuesday 06 March 2007
- Calgo Session (2h00') Krisztian Peters, Patrice Verdier
- Jet-ID Mikko
- MET Lidija
- Electron/Photon-ID Dmitry, Oleksiy
- Tau-ID Peter
- L3 Taus Sohrab
- Jets/Z->ee Patrice
- Tralgo Session (1h45') Ron Lipton, Peter Ratoff, Mike Strauss
- L3 Tracking Ray
- L3 Vertexing Phil Vint
- Vertexing DooKee
- p20 Vertex Certification Monica
- A look at Problematic Tracks Mike H.
- Tracking Performance Mike S.
- Muon-ID Mike Eads, Boris Tuchming
- L3 muons Nirmalya Parua
- Data/MC Performance at High Lumi M. Eads
- muon acceptances at high luminosity B. Tuchming
- B-ID Don Lincoln
- Physics Analyses at High Luminosity
- Results from PhysicsCert Emanuel Strauss
- B_s->mu+mu- search at high lumi Ralf Bernhard
- B Physics Triggering and Physics Channels Penny Kasper
- Forward Muon Yield vs LuminosityValeri Evdokimov
- Z->ee (from xsec) efficiency vs. lum. Jon Hays

Tracking at high luminosity

- DØ Run II central tracker was designed for 200 E30 at 132 ns bunch spacing
- Currently running at 300 E 30 at 396 ns
- Very difficult for tracking at highest luminosities due to high occupancy in tracker

Occupancy vs CFT layer

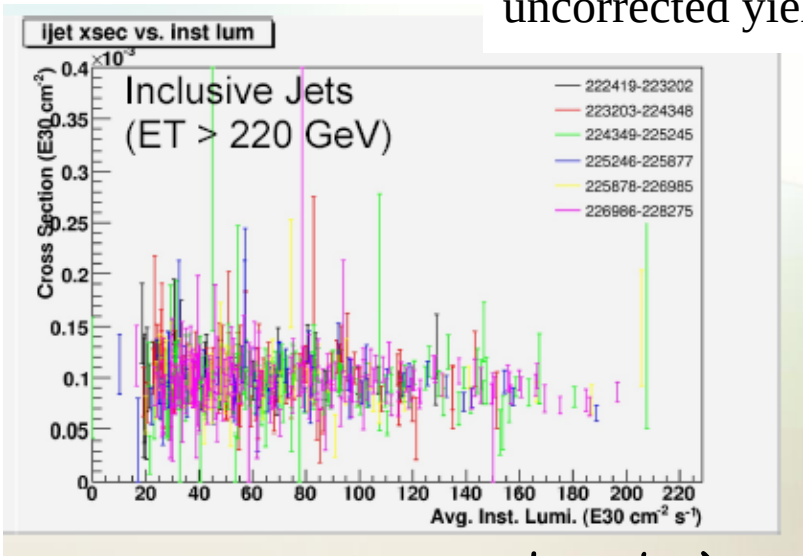
Physics run trigger mix



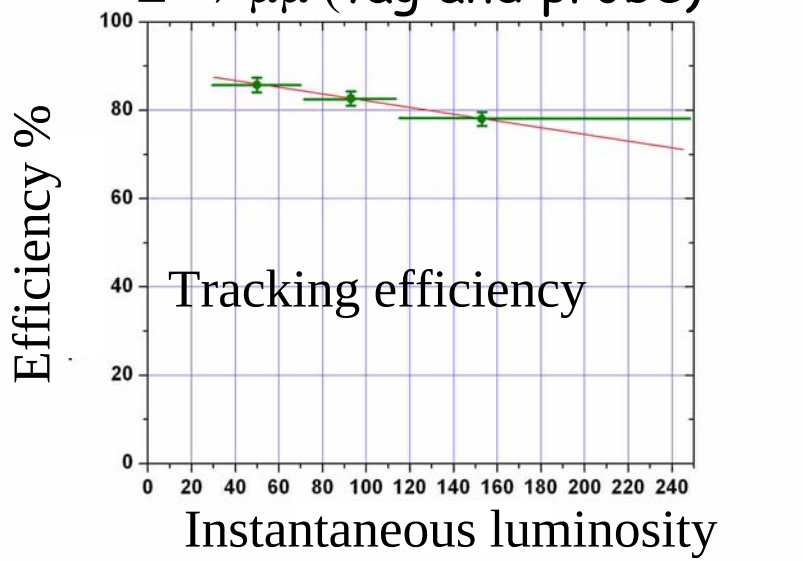
High Luminosity

- Currently in process of performing detailed studies of algorithm performance as a function of instantaneous luminosities
- Jet cross section above 220 GeV shows no instantaneous luminosity effects
- Loss of tracking efficiency observed at high instantaneous luminosity
- Studies of tracking performance at high luminosities are expected to be completed within 2 months.
- This will provide opportunity to optimize accelerator +detector complex through improvements in algorithms and/or store duration, luminosity leveling

Normalized but uncorrected yields

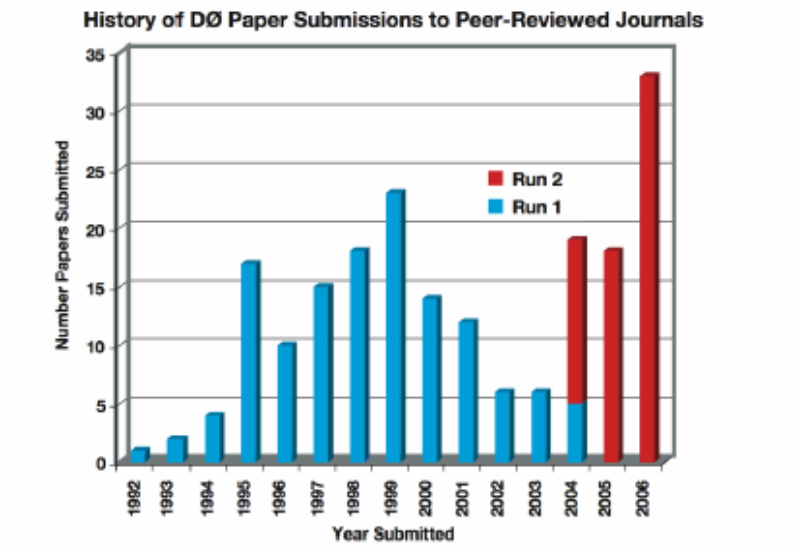


$Z \rightarrow \mu\mu$ (tag and probe)



Getting out the physics

- 2006: Best year ever for DØ publications
- Direct limits on B_s^0 oscillation frequency (PRL **97**, 021802 2006)
  - Most cited experimental particle physics result in 2006 (101 citations and counting)
- First evidence of single top and first direct measurement of $|V_{tb}|$(hep-ex/0612052)



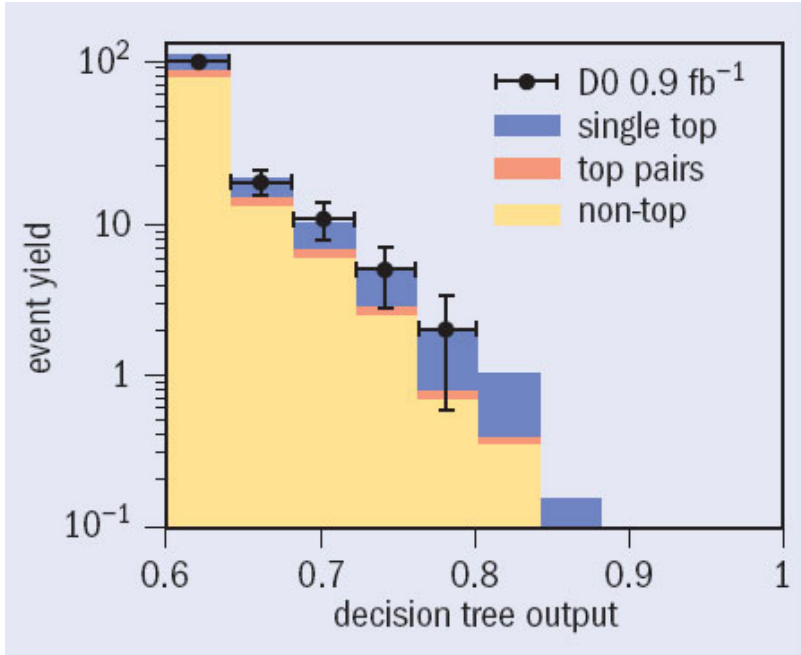
- Publications in
  - QCD
  - EW
  - Higgs
  - New Phenomenon
  - B physics
  - Top

Most physics groups use CAF/CAFe (common analysis format) which allows users to apply efficiencies, b-id efficiency/purity and trigger efficiencies which are centrally produced

Physics groups using common definitions whenever possible

Top physics with 1 fb^{-1}

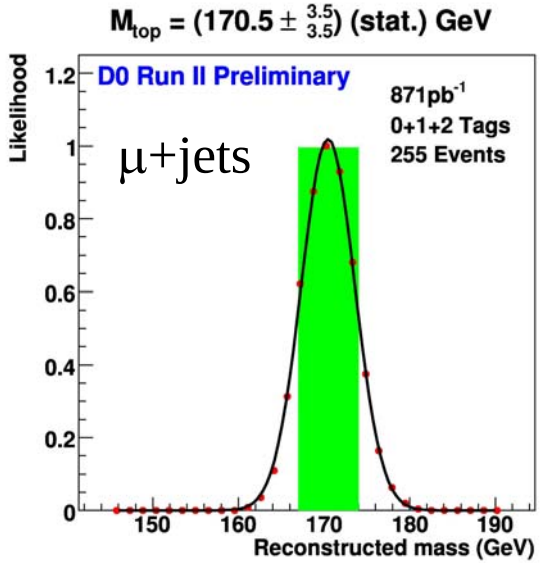
- Search for single top production
- Single top cross section not much smaller than for top quark pair production, but backgrounds are larger
- Three techniques-boosted decision trees, matrix element based likelihood discriminant, and Bayesian neural networks
- First evidence for single top and first direct measurement of $|V_{tb}|$
- Cross section for single top quark production found to be 4.9 ± 1.4 pb ($3.4\ \sigma$ significance)
- $0.68 < |V_{tb}| < 1$ at 95% CL
- Consistent with SM expectations

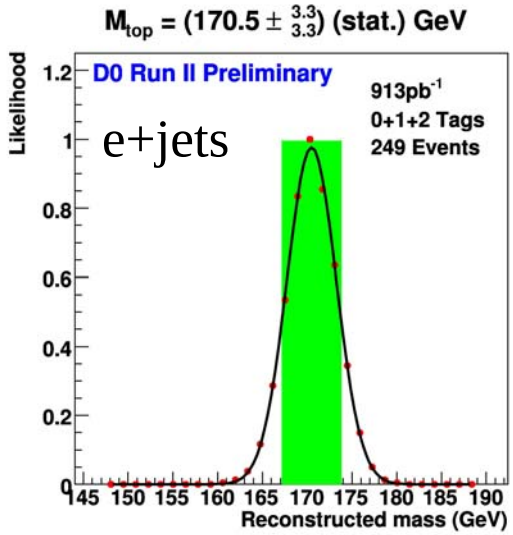


Top Mass with 1 fb^{-1}

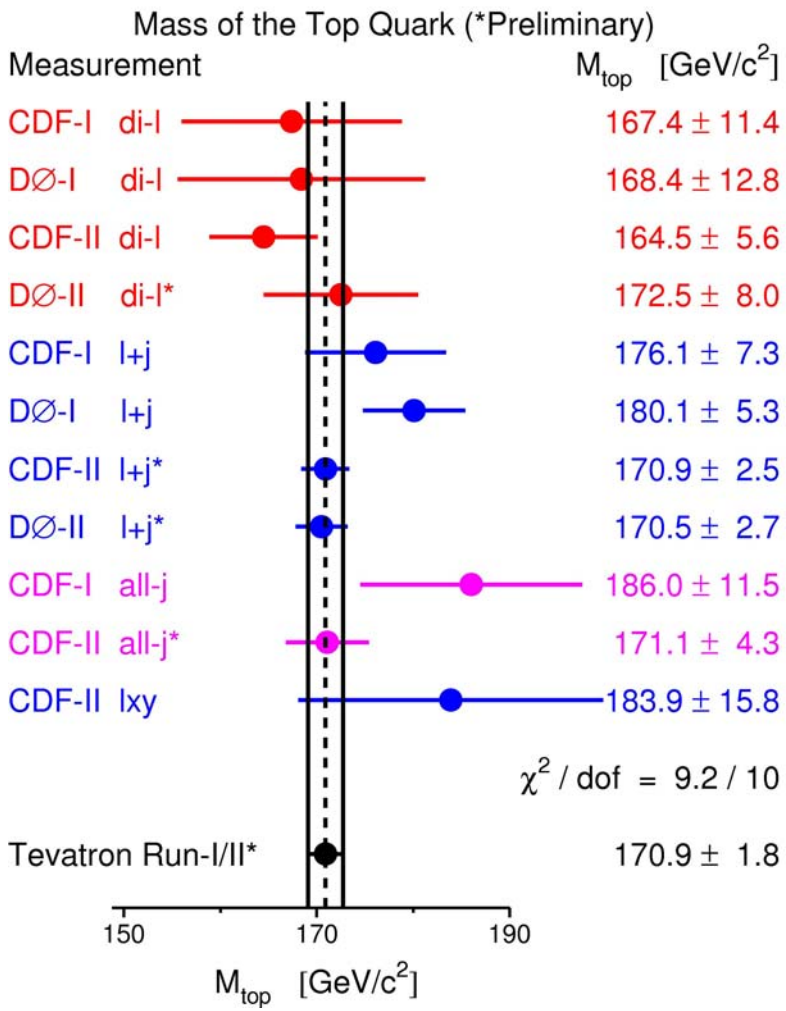
- Top Mass measurement in lepton plus jets using matrix element method
- Purity increased by using neural net b-tagging
- To improve result, in addition to top mass, overall jet energy scale calibration included
- Further significant improvements to jet energy scale expected soon

- M_{top}= 170.5 ±2.4(stat+JES) ± 1.2 (syst) GeV



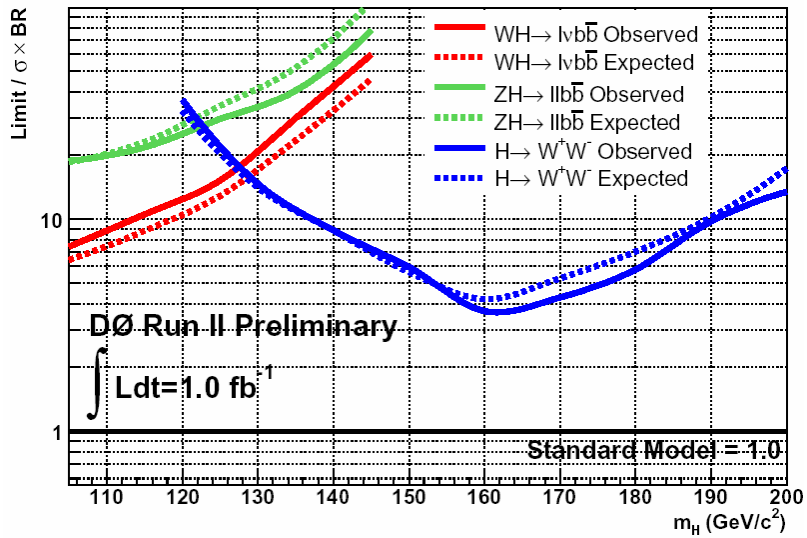


Tevatron Top mass



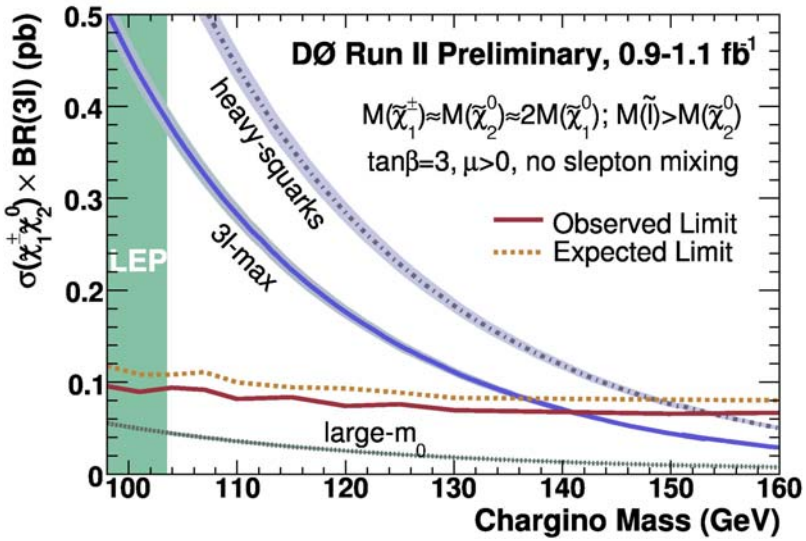
Higgs with 1 fb^{-1}

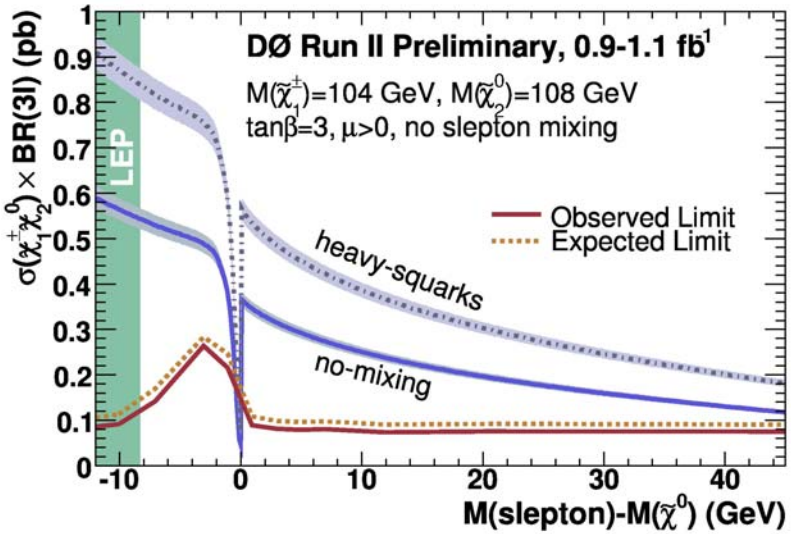
- With 1 fb^{-1} of data, DØ is able to reach Higgs sensitivity only factor of 8(4) at Higgs mass 115 (160) GeV with respect to SM expectations
- In addition to sqrt(L) improvements expected
  - Additional use of Neural nets
  - Neural net b-tagging
  - Layer 0 improvements
  - Improved mass resolution
  - Increased acceptance
- With 2-4 fb^{-1} of data expect to be within striking distance of SM



New Phenomenon with 1 fb^{-1}

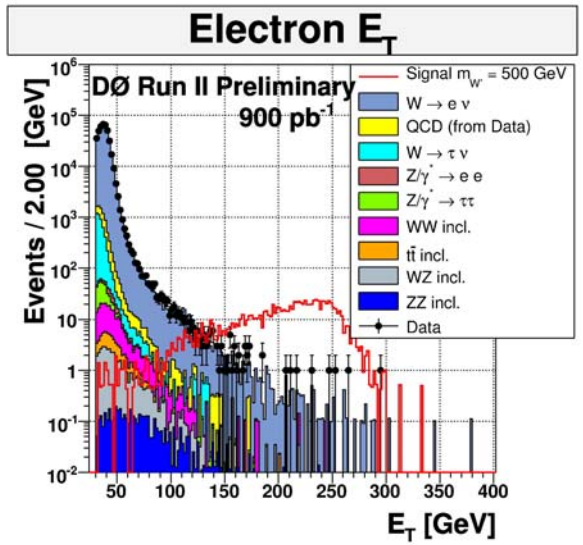
- Search for trilepton decays from associated production of lightest chargino and next to lightest neutralino
- No evidence of supersymmetry and new upper limits on cross section x branching ratio are set
- Currently most stringent limits

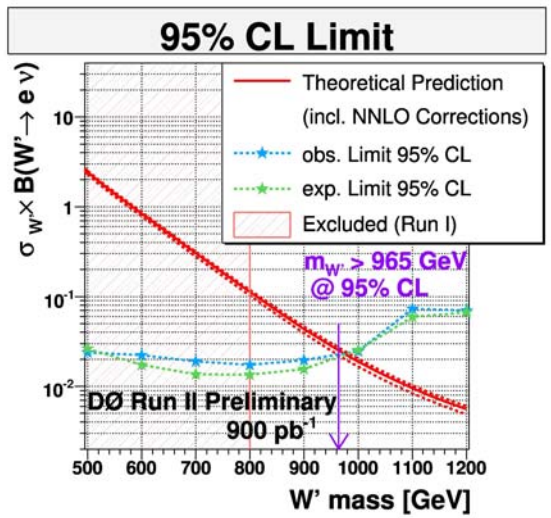




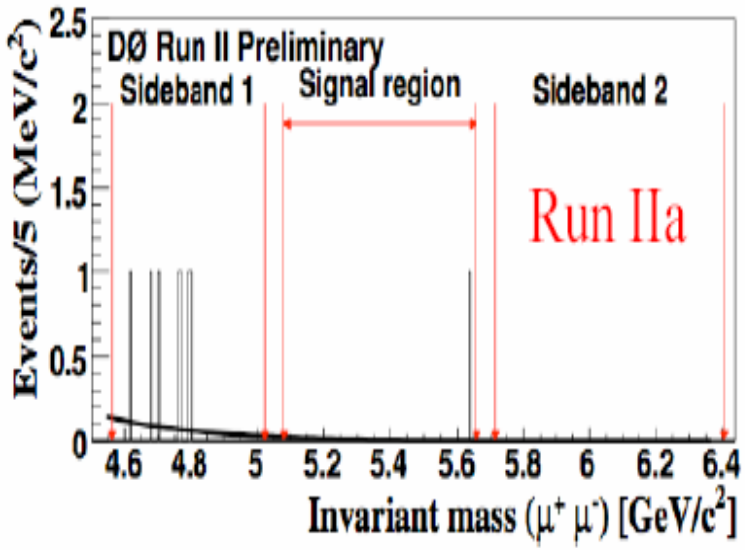
$W' \rightarrow e\nu$ with 1 fb^{-1}

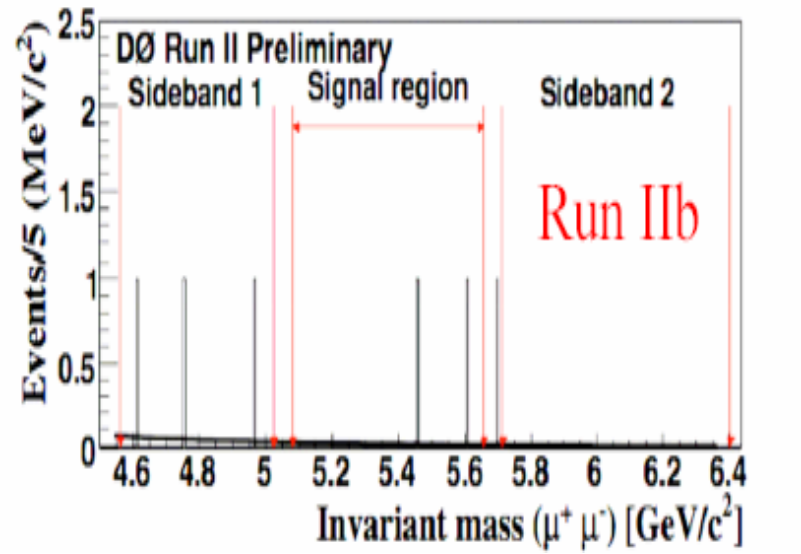
- Some extensions of the SM introduce an additional charged gauge boson W'
- Search for W' decaying to an electron and a neutrino
- No significance excess is seen
- Assume that W' has same couplings to fermions as Standard Model W
- A W' with a mass below 965 GeV can be excluded at the 95% CL





First 2 fb^{-1} results





- $B_s \rightarrow \mu^+\mu^-$
- BF < 7.1 x 10^{-8} at 90% CL
- Currently world's best limit
- B physics can get results out quickly due to the fact that it often doesn't need to wait for calibrations, Jet energy scale, b-id.

Conclusions

- DØ is efficiently collecting data and is able to keep up with reconstruction of data
- DØ is efficiently publishing papers using the large data set collected
- Working to streamline analyses and production to reduce manpower needs and to produce physics results more quickly
- Continue to move toward common tools on the Grid
- Challenges
  - Continue to exploit grid resources in LHC era
  - Find individuals with specific skill sets in future
  - Continue to develop efficient algorithms for high luminosity

Backup Slides

PPD/CD Funding

- Particle Physics Division

	FY06 plan	FY06 actual	FY07 request
DØ operation	1.333 M	1.587 M	1.323 M
Guests and Visitors	0.733 M	0.712 M	1.050 M
Total	2.066M	2.299 M	2.373 M

- Computing Division FY07 request
  - Operation: 0.5M
  - Equipment: 1.25 M (no budget yet)
  - Guest Scientists: (3 people)
- FY08 and FY09: expect to be similar to FY07 for both PPD and CD

PPD/CD Resources

- Computing support
  - 3 FTE for network and windows
  - 9 people from CD system management group (part time)
  - 3.5 FTE for data handling
  - 2 people on databases (part time)
- Future Effort
  - Continue to play significant role in DØ
  - Recently added
    - 3 new application scientists
    - 4 new postdocs
    - 5 postdoc openings(4 in PPD, 1 in CD)
  - Technical support manpower will continue at current level
- Physicists manpower projections
  - 2007: 35 FTE
  - 2008: 28 FTE
  - 2009: 25 FTE
- Strong continued support from PPD/CD through 2009

Security

- CIAC #545339 12/27/2007 Impact:D0, AD, TD
  - Alert:
    - Scanning behavior of several nodes was detected by Fermilab Darknet
  - Status:
    - Four nodes found to be infected with Symantec AV worm, one at D0, two in AD, one in TD. Nodes were actively scanning network for connections on port 2967
  - Root Cause:
    - Numerous nodes were found to be running vulnerable version of Symantec AV. Some not updated due to failure of auto-updates which had not been detected. Some running vulnerable versions because Symantec incorrectly reported one version as safe. Some running vulnerable versions due to inconsistency in way Symantec reports installed versions.
- No data lost nor any impact on D0 Online Major Application

Security

- CIAC # 546321 01/09/2007 Impact: D0
  - Alert:
    - Phone call from DHS reporting that Fermilab node was seen participating in IRC chat room monitored by DHS
  - Status:
    - Machine was found to have IRC daemon installed. User had attempted to open "postcard" sent to him from unknown source. Postcard installed IRC daemon.
  - Root Cause:
    - Pilot error
- No data lost nor any impact on D0 Online Major Application

- "Reportable" Incident
  - Incidents are reported according to the Incident Prevention Warning and Response (IPWAR) requirements in DOE directive 205.1.1. It is the responsibility of the Fermilab Computer Security Coordinator (FCSC) to see that reporting is done in compliance with the current IPWAR manual.
  - See http://www.directives.doe.gov Search for directive 205.1.1

Algorithms manpower needs

Algorithms	FY07	FY09
Tracking	2.9	1.9
Muon Reconstruction	2.5	1.0
Calorimetry, e and γ	7.5	2.3
Taus	1.4	0.6
Jet Energy scale	6.8	1.2
b-tagging	4.0	1.5
Trigger	17.4	4.0
Simulation	4.0	3.5
Luminosity	1.5	0.5
High level data handling	6.5	4.5
Infrastructure	0.4	0.4
Total	54.9	21.4

Computing manpower needs

Position	Development	Steady State
Management	4.3	4.3
Central Farm	0	0
MC production	1.0	0.5
Reprocessing	1.8	1.8
Sys. Administation	3.0	3.0
Data Bases	2.5	1.2
Builds and core support	1.0	0.5
Data Handling	1.0	0.8
Distributed computing	16.0	11.5
Job Control/tools	1.5	1.5
Subtotal	32.1	25.1
CD central contribution	13.9 + 8 shared	12.9 + 6 shared
Total	46.0 + 8 shared	38.0 + 6 shared

Physics Manpower needs

Physics	FY07	FY09
Bs: mixing,rare decay,lifetime	15.5	11.5
EW: W mass	13.5	8.0
Higgs: SM and Susy	17.5	16.0
New Phenomenon : Trilepton,squarks,gluinos, stop and sbottom,LED, Z'	9.0	6.0
Top: Mass and single top production	16.8	12.0
Core physics management	7.0	6.0
Tevatron Combination	3.0	3.0
God Parents/Editorial boards	4.0	4.0
Total	86.3	66.5